

⑨ 日本国特許庁(JP)

⑩ 特許出願公開

⑪ 公開特許公報(A)

昭62-24481

⑫ Int. Cl.⁴

G 11 B 20/18
G 06 F 3/06

識別記号

1 0 2
3 0 5

庁内整理番号

6733-5D
6711-5B

⑬ 公開 昭和62年(1987)2月2日

審査請求 未請求 発明の数 2 (全11頁)

⑭ 発明の名称 デジタル・データ・ワードの格納兼検索システム

⑮ 特 願 昭61-105782

⑯ 出 願 昭61(1986)5月8日

優先権主張 ⑰ 1985年5月8日 ⑱ 米国(US) ⑲ 732353

⑳ 発 明 者 ダブリュー・ダニエル・ヒリス アメリカ合衆国マサチューセッツ州ケンブリッジ, グリーン・ストリート 548
㉑ 発 明 者 クレメント・ケイ・リウ アメリカ合衆国マサチューセッツ州ブライトン, クレイモス・ロード 41
㉒ 出 願 人 シンキング・マシーンズ・コーポレーション アメリカ合衆国マサチューセッツ州02142, ケンブリッジ, ファースト・ストリート 245
㉓ 代 理 人 弁理士 湯 浅 恭 三 外5名

明 細 書

1. [発 明 の 名 称]

デジタル・データ・ワードの格納兼
検索システム

2. [特 許 請 求 の 範 囲]

(1) デジタル・データ・ワードの格納および検索を行ない、検索されたデータ・ワードにおけるエラーを訂正するシステムにおいて、

前記各データ・ワードを機械的に駆動される媒体上に格納するためのデータ記憶装置と、

前記データ・ワードに対するエラー訂正ビットを生成するためのエラー訂正回路と、

機械的に駆動される媒体における前記データ・ワードとは独立的に前記訂正ビットを格納するための訂正ビット記憶装置とを設け、

前記エラー訂正回路は、前記エラー訂正ビットを用いて前記の検索されたデータ・ワードにおけるエラーを訂正するように構成されることを特徴とするシステム。

(2) 前記データ記憶装置が、前記各データ・ワ

ードの異なる部分をそれぞれ格納するためのディスク記憶装置を含み、

前記訂正ビット記憶装置が、少なくとも1つの訂正ビット記憶装置を含むことを特徴とする特許請求の範囲第1項記載のシステム。

(3) 前記エラー訂正回路が、どれか1つの前記記憶装置の障害を許容するように前記データ・ワードを訂正するよう構成されることを特徴とする特許請求の範囲第2項記載のシステム。

(4) 複数の訂正ビット記憶装置が設けられ、

種々の前記エラー訂正ビットが異なる前記訂正ビット記憶装置に格納されることを特徴とする特許請求の範囲第2項記載のシステム。

(5) ある前記データ・ワードに対する2つの前記エラー訂正ビットが1つの前記訂正ビット記憶装置に格納されないことを特徴とする特許請求の範囲第2項記載のシステム。

(6) 前記エラー訂正回路が、

一時に前記各データ・ワードにおける全ビットより少ないビットに対して1組の前記エラー訂正

ビットを生じるエラー訂正装置と、

連続するグループにおける前記データ・ワードのビットを前記エラー訂正装置に対して経路指定を行なつて複数の組の前記エラー訂正ビットを生成する経路指定回路とを含むことを特徴とする特許請求の範囲第2項記載のシステム。

(7) 前記エラー訂正回路が、

その各々が前記各データ・ワードにおける全ビットより少ないグループに対して1組のエラー訂正ビットを生成する複数のエラー訂正装置と、

前記エラー訂正装置に対して並列に前記各データ・ワードにおけるビットのグループの経路指定を行なつて複数の組の前記データ・ビットを生成する経路指定回路とを含むことを特徴とする特許請求の範囲第2項記載のシステム。

(8) 前記各データ・ワードが64ビットを有し、前記各エラー訂正装置が前記各データ・ワードの32ビットの各グループに対して7つのエラー訂正ビットを生成することを特徴とする特許請求の範囲第6項または第7項に記載のシステム。

他方の前記記憶装置に格納された前記データ・ワードの部分が前記バスに対して送出される用意がある時との間の前記第1の遅延時間とは異なる第2の遅延時間を有し、前記バスと前記記憶装置との間に接続されて、前記記憶装置にそれからの連続する前記各ワードの前記部分を読出させることにより、前記データ・プロセッサからの前記要求に回答するアダプタを設けることを特徴とするシステム。

(11) 前記アダプタが更に、前記ワードの全ての部分が送出される用意ができるまで前記バスに対するどの前記部分の送出を遅延させるように構成されることを特徴とする特許請求の範囲第10項記載のシステム。

(12) 前記各記憶装置と前記アダプタとの間に個々のバスを更に設け、

前記各記憶装置は、前記記憶装置に格納された前記データの前記部分が送出の用意がある時、その対応する個々のバス上にドレーン指令を前記アダプタに対して送出するように構成され、

(9) 前記データ・ワードが、格納および検索のための並列データ・バスを介して順次送出され、

前記バスと前記記憶装置間に接続されて、該バスと記憶装置間のデータ・ワードの経路指定を行なうアダプタを更に設けることを特徴とする特許請求の範囲第2項記載のシステム。

(10) デジタル・データ・ワードを格納し、前記バスに対して接続されたデータ・プロセッサにより要求される時、並列データ・バスに対して順次デジタル・データ・ワードを送出するシステムにおいて、

各々が機械的に駆動される媒体と該媒体の読出しを行なうヘッドとを備えた記憶装置を設け、前記各データ・ワードの異なる部分が異なる前記記憶装置において格納され、一方の前記記憶装置は、特定の前記データ・ワードが要求される時と前記記憶装置に格納された特定の前記データ・ワードが前記バスに対して送出される用意がある時との間の第1の遅延時間を有し、他方の前記記憶装置は、特定の前記データ・ワードが要求される時と

前記アダプタは、全ての前記の個々のバス上の前記ドレーン指令の発生を監視するように構成され、また全ての前記記憶装置からの前記ドレーン指令が現われる時、前記記憶装置からの前記データ・ワードの前記部分の全てをドレーンし、かつ前記並列データ・バスに対してこれら部分を同時に送出するように構成されることを特徴とする特許請求の範囲第9項または第10項に記載のシステム。

(13) 前記アダプタが、前記各データ・ワードに対して並列エラー訂正ビットを生成するようになっており、

前記エラー訂正ビットの少なくとも1つが、前記エラー訂正ビットと関連するデータ・ワードの前記部分を保持しない記憶装置において格納されることを特徴とする特許請求の範囲第10項記載のシステム。

(14) 前記アダプタは更に、関連する前記データ・ワード部分が読出される時点と対応する時点において前記エラー訂正ビットを読出させるように構

成され、

前記アダプタは、前記エラー訂正ビットに基く前記データ・ワードにおけるビットのエラーを訂正するためのエラー訂正回路を含むことを特徴とする特許請求の範囲第13項記載のシステム。

09 前記エラー訂正ビットの異なるグループが異なる前記記憶装置において格納されることを特徴とする特許請求の範囲第13項記載のシステム。

06 前記各データ・ワードに対する前記各エラー訂正ビットが個々の前記記憶装置に格納されることを特徴とする特許請求の範囲第15項記載のシステム。

07 更に、前記データ・プロセッサにより指令される時、前記データ・ワードを格納するように構成され、

前記各記憶装置は更に、異なる前記記憶装置に格納されつつある前記各データ・ワードの異なる部分を前記媒体上に書き込むためのヘッドを含み、異なる前記記憶装置は、特定の前記ワードが格納するため送出される時と前記記憶装置がその各々

前記アダプタは、どの前記制御装置におけるバッファも空の状態にならないように保証するに充分な速度で連続する前記データ・ワード部分の前記記憶装置からの送出を行なうように構成されることを特徴とする特許請求の範囲第9項または第10項に記載のシステム。

02 異なる前記記憶装置に格納された前記各データ・ワードの部分が等しいことを特徴とする特許請求の範囲第9項または第10項に記載のシステム。

01 前記部分が前記データ・ワードの各2ビットであることを特徴とする特許請求の範囲第20項記載のシステム。

02 個々の前記各バスがSCSIプロトコルに準換することを特徴とする特許請求の範囲第10項記載のシステム。

03 前記各データ・ワードに32ビット以上が存在することを特徴とする特許請求の範囲第1項または第10項に記載のシステム。

20 前記各データ・ワードに256ビットが存在

の前記部分を格納する用意がある時との間の異なる遅延時間を有し、

前記アダプタが更に、全ての前記記憶装置が格納する用意ができるまでどれかの前記部分の前記記憶装置に対する送出を遅延させることにより、前記データ・プロセッサからの前記指令に応答するように構成されることを特徴とする特許請求の範囲第10項記載のシステム。

08 前記各記憶装置が、前記媒体に対して書き込まれるべきビットを一時的に保持するためのバッファを備えた制御装置を含み、

前記アダプタは、どの前記制御装置におけるバッファも空の状態にならないように保証するに充分な速度で連続する前記データ・ワード部分の前記記憶装置に対する送出を行なうように構成されることを特徴とする特許請求の範囲第17項記載のシステム。

09 前記各記憶装置が、前記媒体から読出されて送出される用意があるビットを一時的に保持するためのバッファを備えた制御装置を含み、

することを特徴とする特許請求の範囲第1項または第10項に記載のシステム。

05 前記データ・ワードが、前記バスの回線当り少なくとも4メガビットの速度で前記並列データ・バスに対して送出されることを特徴とする特許請求の範囲第8項または第10項に記載のシステム。

04 前記各記憶装置は磁気ディスク駆動装置を含むことを特徴とする特許請求の範囲第2項または第10項に記載のシステム。

3. [発明の詳細な説明]

[発明が属する技術分野]

本発明は、磁気ディスクの如き機械的に駆動される記憶媒体に対しデジタル・データを格納する装置に関する。

[従来の技術およびその問題点]

典型的な磁気ディスク駆動装置においては、いくつかの磁気ハード・ディスクが単一の回転軸上に取付けられている。各ディスク面上の記憶スペースは、同心円状のトラックに構成されている。

全てのディスク面における対応するトラックは、仮想のシリンダを形成する。各ディスク面は、読出し／書き込みヘッドにより機能させられる。全ての読出し／書き込みヘッドは、選択されるどのシリンダに対しても一緒に運動させることができ、また従つてこのシリンダを構成するトラック上にデータを格納しあるいはその上のデータの検索ができる。

通常、磁気ディスク駆動装置は直列バスによりディスク制御装置に対して接続されており、この制御装置は、就中、読出し／書き込みヘッドを適正なトラックに対して運動させ、ディスクに格納されようとするかあるいはディスクから検索されたデータのデータをバッファし、データが適正なアドレスに格納されあるいはこのアドレスで検索されることを保証し、かつデータに関する逐次のエラーの訂正および検出を行なう。典型的には、このディスク制御装置は、その入出力ポートが例えば32ビット幅であるコンピュータに対して接続された並列の入出力バスを提供する。このバス上に

るための少なくとも1つの訂正ビット記憶装置がある。エラーは、どれか1つの記憶装置の障害に対し格度を有するように訂正される。あるデータ・ワードに対するエラー訂正ビットは様々な異なる訂正ビット記憶装置に格納されている(ある1つのデータ・ワードに対して2つのエラー訂正ビットが1つの記憶装置に格納されないことが望ましい)。1組のエラー訂正ビットが、一様に1つのデータ・ワードにおける全てのビットよりも少数のグループ毎に生成され、データ・ビットのグループは次々に処理されて連続する組のエラー訂正ビットを生じる。あるいはまた、異なるグループのデータ・ワードのビットが同時に並列に処理されて、エラー訂正ビットの組を生じる。例えば、各データ・ワードは、それぞれ32ビットの2つのグループを構成する64ビットを持つことができ、1組の7つの訂正ビットを32ビットの各グループ毎に生成することができる。データ・ワードは、格納および検索のため並列のデータ・バスを介して順次送出され、バスと記憶装置間に接続

送られる32ビットのワードは、ディスク制御装置を介して送られ、順次ディスク制御装置に対してロードされあるいはこれから検索される。ディスク制御装置とディスク駆動装置間の直列バスの処理速度は、おおよそコンピュータとディスク制御装置間の並列のI/Oバスの処理速度と整合されている。

全ての利用可能な記憶スペースを拡張するため、多重ディスク駆動装置をデジー・チェーンまたは星状回路形態で単一の制御装置に対してリンクすることができる。

[問題を解決する手段]

本発明の全般的な1つの特徴は、データ・ワードに対するエラー訂正ビットを生成すること、およびこのデータ・ワードとは独立的に機械的に駆動される別個の媒体上に訂正ビットを格納することにある。

本発明の望ましい実施態様は下記の特徴を含んでいる。各データ・ワードの異なる部分は異なるデータ記憶装置に格納され、訂正ビットを格納す

られてこのバスと記憶装置間にデータ・ワードを送るアダプタが提供される。

本発明の別の目的は、異なる非同期的な機械的に駆動される記憶装置における並列バス上に現われる各データ・ワードの異なる部分を格納することにある。

本発明の望ましい実施態様は以下の特徴を含む。1つのデータ・ワードの全ての部分が送られる用意ができるまで、このデータ・ワードのどの部分のバスに対する送出の遅れによつてデータ・ワードの検索が行なわれる。各記憶装置は、要求されたデータ・ワードのその部分を送出する用意ができる時ドレーン指令を送り、全ての記憶装置がドレーン指令を出した時のみ記憶装置はドレーンされる。並列のエラー訂正ビットが各データ・ワード毎に生成され、このエラー訂正ビットの少なくとも1つが関連するデータ・ワードのどの部分も保持しない記憶装置に対して格納される。エラー訂正ビットは、データ・ワードの読出しと訂正を行なう時点において読出され、データ・ワードに

おける訂正する1つのビットのエラーに対して使用される。データ・ワードを格納する時、データ・ワードの異なる部分は、これらが全て格納の用意ができるまで、異なる記憶装置に対しては送られることはない。各記憶装置は1つのバッファを備え、読出されるべき連続データ・ワードがバッファが充満状態とならないことを保証するに十分な速度で要求され、書き込まれるべき連続するワードがビットが空にならないことを保証するに十分な速度で送出される。各データ・ワードの等しい数のビット(例えば、2ビット)が異なる記憶装置において格納される。1つの記憶装置を供する各個のバスはSCSIプロトコルに準拠している。各データ・ワードには32ビット以上(例えば、256ビット)が存在する。これらのデータ・ワードは、バス回線毎に少なくとも4メガビットの速度で並列バスに対して送出される。各記憶装置は磁気ディスク駆動装置である。

本発明は、非常に広範囲の高速の並列I/Oバスに供することができる非常に大きな処理速度の

14を介して、このKBUS上で受取ったデータを格納しかつこのKBUSに対して検索されたデータを送出するための対応する1つの高速ディスク・サブシステム16に対して接続されている。各KBUSは毎秒32メガバイト(バイト当り8ビット)の処理速度を有し、データは毎秒128メガバイトの総処理速度で4つの全てKBUS上に同期的に送ることができる。

第2図によれば、各ディスク・サブシステム16はKBUSの1つに接続されたバス・アダプタ18を有する。このバス・アダプタ18はまた、各々8ビット幅の39の異なる並列バス(SCSIバス)20を介して、39個の異なるディスク記憶装置22に対して接続されている。各ディスク記憶装置22は、直列バス(ESDIバス)26を介してディスク制御装置28(例えば、Emulex社の販売するチャンピオン・モデル)に対して接続されたディスク駆動装置24(例えば、Control Data社のモデル9415)を有し、前記制御装置は更に前記ディスク記憶装置に供されるSCSI

記憶システムを提供するものである。標準的な市販のディスク駆動装置および制御装置は、異なるディスク駆動装置の主軸の回転運動と同期する必要なしに使用することができる。エラー訂正ビットおよびデータ・ビットを独立的に格納することにより、このシステムはどれかのディスク駆動装置全体の故障に対する裕度を有する。標準的なSCSIプロトコルの指令を用いてシステムの全体的な動作を同期させる。

本発明の他の利点および特徴については、以下の望ましい実施態様の記述および頭書の特許請求の範囲から明らかになるであろう。

[実施例]

(構造および作用)

第1図によれば、(本願と同じ譲受人に譲渡され参考のため本文に引用される1983年5月31日出願の米国特許出願第499,474号に記載された種類の)コンピュータ10が4つの同じ入出力(I/O)ポート12を備えている。各I/Oポート12は、64ビット幅の並列バス(KBUS)

バスに対して接続されている。各SCSIバスは、参考のため本文に引用される米国商務省、規格標準局の規格刊行物第X3T9号、改定第14B号(1984年11月6日)に規定された小型コンピュータ・システムのインターフェース規格に従って、データおよび指令(指令記述子ブロッカーCDBの形態における)を送る。

39個のディスク記憶装置および39個のSCSIバスは2つのグループに構成されている。ディスク記憶装置およびバスの内の32個(D1乃至D32で示される)は、コンピュータ10からのデータの格納および送出的ため使用される。残りの7個のディスク記憶装置およびバス(E1乃至E7で示される)は、以下に述べる方法でデータ・ビットから送られるエラー訂正ビットの格納および送出的ため使用され、エラーの訂正のために使用される。これらが保持する情報のタイプにおけるこの区別を除いて、39個の全てのディスク記憶装置およびSCSIバスは同じものである。

各ESDIは、例えば、参考のため本文に引用

されるMagnetic Peripherals社の刊行物第77738076-D号(1984年)において規定された強化型小型装置インターフェースに従ってデータおよび指令を逐次を送る。

第3図によれば、各バス・アダプタ18においては、KBUSの64回線はそれぞれ32回線の2組に分割され、この2組がそれぞれ2つのエラー検出兼訂正装置32(それぞれ例えば1対のIntel社の8206型またはAMD社の2960型チップからなる)の両方向のWD/CD入力(データまたは訂正データの書き込み)側に対して接続されている。各WD/CD入力に対して接続された32の回線もまた、それぞれ32個の8ビットのシフトレジスタ34の入力側に対して接続されている。このように、各シフトレジスタはその2ビットの入力が2組の32の回線の各々からの1回線に対して接続され、各シフトレジスタは一時に2ビットをシフトするように構成されている。

各装置32のWD/CD入力に対して加えられる32のデータ・ビットの各組毎に、本装置が7つ

の並列のエラー訂正ビットを生じ、これがこれらのビットの以降の検索中に生じる関連した32のデータ・ビットのどれかにおける単独ビットのエラーを訂正するに十分な情報を提供する。

各装置32の7つのエラー訂正ビットが7つのエラー訂正ビット回線を介してSC出力側から送出され、このビット回線はレジスタ34の如き7つの8ビットのシフトレジスタ36の入力側にそれぞれ接続されている。

KBUSの1つの転送サイクルが経過した後、各シフトレジスタ34はKBUSの回線の内2つに現われるデータ・ビットと対応する2つのデータ・ビットを保有し、また各シフトレジスタ36は対応するエラー訂正ビットの内の2つを保有する。KBUSの4転送サイクルが経過した後、各シフトレジスタ34は8つのデータ・ビット(1データ・バイト)を保有し、各シフトレジスタ36は8つのエラー訂正ビット(1エラー訂正バイト)を保有する。その時、シフトレジスタ34、36におけるバイトはそれぞれSCSIバスD₁~D₃₂

およびE₁~E₇に対してアンロードされる。次いで、KBUSの次の4つの転送サイクル毎にこのプロセスが、...というように繰返される。シフトレジスタ34、36のローディングおよびアンローディングは、KBUS上で毎秒32メガバイト(回線毎に4メガバイト)の頻度で現われるデータがSCSIバス当たり毎秒1.5メガバイトの速度で8ビット幅の32のSCSIバスに対して送られる。反対に、SCSIバス当たり毎秒1.5メガバイトで現われるデータは毎秒32メガバイトでKBUSに対して送られる。

データがSCSIバスの各バイトの転送サイクルにおいてディスク駆動装置24から検索されると、1つの8ビット・バイトが平行に各シフトレジスタ34、36に対してロードされる。次に、全てのシフトレジスタ34が一様に平行的にそのエラー訂正バイトの第1と第2のビットをEDC装置32のRD(データ読出し)入力に対してアンロードし、全てのシフトレジスタ36が平行的にそのデータ・バイトの第1と第2のビットを

EDC装置32のCB(検査ビット)入力に対してアンロードする。EDC装置32は、そのCB入力を介して受取られたエラー訂正ビットを用いてそのRD入力で受取ったデータ・ビットに現われる単独ビット・エラーを調べてこれを訂正する。その時、EDC装置32はそのWD/CD出力側における64の訂正されたデータ・ビットをKBUSに対して送出する。このプロセスは、シフトレジスタ34、36に保持されたバイトの全ての8ビットがアンロードされるまで繰返される。次いで、シフトレジスタはSCSIバスからの新たなバイトで再びロードされる。

各ディスク記憶装置28(第2図)は、下記の如くディスク駆動装置24上のデータの格納および検索を制御するように構成されている。

格納操作を開始するため、コンピュータ28はKBUS上に適当な指令記述ブロック(CDB)を送つて、特定数のバイト(ワード・カウント)が特定のディスク・アドレスから始めて格納されるべきことを表示する。このCDBは、バス・ア

アダプタ18によりSCSIバスを介してディスク記憶装置28に対し送られる。制御装置は、必要に応じて、ESDIバス上の探索指令をディスク駆動装置に対して送り、これをして読出し／書き込みヘッドをCDBにそのアドレスが指示された格納場所を含むシリンダへ移動させる。制御装置28は次に、格納されるべきデータがSCSIバス上へ送出されることを要求する。この要求は、再びKBUSを介してコンピュータへ戻される。次に、このデータは、(前述の如く)KBUS上をバス・アダプタを介してSCSIバスに対し送られ、データ・ビットが受取られる時、これらビットは制御装置における14Kバイトのバッファに格納される。各トラックは約10Kバイトの容量を有するため、前記バッファは一時に全トラック相当以上のデータを保持することができる。ディスク記憶装置28は、バッファが充満状態になる時(ワード・カウントがバッファ容量より大きいかあるいはこれと等しい場合)、あるいはさもなければバッファにおけるバイト数がワード・カウン

タは、ESDIバス上を順次制御装置の直列バッファに対してロードされる。バッファがこのトラックにおけるバイト数と等しいバイト数を有すると直ちに、制御装置はSCSIバス上をコンピュータ10に対して有を送出してバッファからデータを取得する。このデータがアンロードされつつある間、制御装置はESDIバス上でディスク駆動装置からのデータを受入れて、これをそのバッファに対してロードし続ける。このプロセスは、ワード・カウントと等しいバイト数となるまで続行する。

制御装置に対する全てのSCSIバスを介するデータの格納および検索は平行的に同時に生じる。即ち、バス・アダプタの観点からは、領域種別するデータが単一の操作で同時に全てのSCSIバス上へ送出される。異なるディスク駆動装置24は一般に相互に位相がずれているため、異なるディスク駆動装置における対応するトラックの場所が同時にその各々の読出し／書き込みヘッドの下に現われることが多い。従つて、異なる制御装置が

トと等しければ直ちに、ESDIバスに対してそのバッファのアンロードを順次開始することになる。その時、データはトラック上の指定されたアドレスから始つて格納される。もしワード・カウントがバッファ容量より大きければ、バッファが8Kバイトのデータしかバッファに残らないレベルまでドレインされるや否や、制御装置がコンピュータ10に対して別のデータを戻すための要求を送出し、制御装置における古いデータがディスク駆動装置24に対してアンロードされると同時に別のデータがバッファに対してロードされる。このプロセスは、ワード・カウントに達するまで続行する。

同様に、検索操作を開始するため、コンピュータは指定されたバイト数が指定されたディスクのアドレスから始めて検索されるべきことを表示するCDBを送出する。必要に応じて、制御装置は最初にディスク駆動装置の読出し／書き込みヘッドをして適正なトラックを探索させ、指定されたアドレスから始めてデータを読出させる。このデー

格納指令を発することができる時点における差、および異なる制御装置が検索されたデータを送出する用意ができる時点における差が存在することになる。異なる制御装置の位相のずれる可能性のある操作は、以下の方法でバス・アダプタに関する同期的な出し入れ伝送の必要と妥協させられる。

第4A図においては、データが格納されるべき時、コンピュータ10が最初にバス・アダプタに対して特定のシリンダを指定する探索指令(60a)を送出する。このバス・アダプタは、次に、全ての制御装置62に対するSCSIバスを介して適当なCDBを送ることにより探索指令を同時通報して、これらの全てのCDBをして同じシリンダを探索させる。制御装置が独立的にこの探索指令を実行中、コンピュータは他の関連のない操作を行なうことができる。

全ての制御装置が探索指令を実行したことを保証する十分な期間(64)の経過後、コンピュータ10は書き込み指令(シリンダ境界が交差しないように関連するディスク・アドレスおよびワード・

カウントを含む)をバス・アダプタに対して送出する(66)。このバス・アダプタは、適当なCDBをSCSIバスを介して全ての制御装置へ送ることによって全ての制御装置に対する書き込み指令(68)を同時通報する。制御装置は全て、この時、格納されるべきデータに対するバス・アダプタに対しSCSIバス上の一連の要求(70)をバイト単位で開始する。異なる制御装置からのデータは、一般に同時には現れない。バス・アダプタは、全てのSCSIバスに対して接続され要求が現れるとこれを検出する論理回路48を有する。論理回路48はまた各シフトレジスタ34, 36の制御入力に対して制御出力が接続され、SCSIバスを介してシフトレジスタのローディングおよびアンローディングを制御する。

全ての制御装置がその要求を発した時、バス・アダプタは論理回路48から全てのシフトレジスタ34, 36に対してアンロード信号を送出することにより全ての要求を同時に満たして、これらシフトレジスタに同じ時点において全てのSISC

ものである制御装置におけるバッファが8Kバイトのレベルにドレーンされるまで待機する。その間、データを更に要求する最初のものであつた制御装置におけるバッファは、8Kバイト・レベルより低くなるよう空の状態になつている。その後、バス・アダプタは、データ・バイトを最初の要求を行なう制御装置が要求すると直ちにデータ・バイトを送出し、これにより最初に要求する制御装置におけるバッファが決して完全に空にならないことを保証する。このプロセスは、各制御装置に対して送られるデータ・バイト数がワード・カウントに達するまで継続する。

図4B図においては、データが検索されるべき時、コンピュータ10は読出し指令(82)が後に続く探索指令(80)を送出し、その両者はバス・アダプタによつてデータ格納操作と類似の方法で制御装置に対して同時通報される(84, 86)。各制御装置は、探索指令(88)を実行し、またディスク駆動装置にディスクからのバイトを読出させることにより読出し指令の実行を開始し、これら

バス上にそのバイトをアンロードさせる(72)。各制御装置は、そのバッファが一杯の状態になるまで(あるいは、そのバッファにおけるバイト数がワード・カウントと等しくなるまで)、データをバイト単位で要求し続けることになる。各制御装置は、その時、そのバッファからのデータをディスクに対して書き込みを開始する(74)。

異なる制御装置は、その各々のディスク駆動装置の軸位置間の位相差に基く非同期時点において書き込み操作を実行することになる。ワード・カウントがバッファ容量を超える場合には、そのバッファに残るバイト数が8Kバイトより下ると直ちに、各制御装置はバス・アダプタから更にデータを要求することになる。これら制御装置のこれ以上のデータに対する要求は、バス・アダプタに非同期的に到達することになる。バス・アダプタは、その論理回路48を介して、より以上のバイトを送ることによりこれら要求を満たす前に全ての要求が受取られるまで待機することになる。実際に、バス・アダプタは、データを更に要求する最後の

バイトを制御装置のバッファにロードされるようESDIバス上で順次送出する。異なる制御装置は、一般に、データ読出しを非同期的に実行することになる。セクターの読出しデータ相当分がバッファに対して入れられた時、制御装置はバッファをドレーンさせるよう要求するが、データ読出しおよびバッファの充填を継続することになる。バス・アダプタの論理回路が全ての制御装置がそのバッファをドレーンさせることを待機中であること検出すると、この論理回路はこれらバッファのドレーン操作を開始する。その後、バス・アダプタは、ドレーンを要求する最初のものであつた制御装置により要求される速度で制御装置のバッファからバイトをドレーンし続けることになり、これにより最初のバッファが決して一杯の状態にならないことを保証する。読出し操作は、各ディスク駆動装置から読出されたバイト数がワード・カウントと等しくなるまで継続する。

各高速ディスク・サブシステムにおけるディスク記憶装置22のどれかが故障すると、EDC接

置は故障した装置から受取つた誤りビットを訂正することが可能である。従つて、ディスク・サブシステムの操作を妨げることなく、故障した装置全体を正常な装置で交換することができる。

本発明は、非常に広い範囲の高速の並列I/Oバスを供することが可能な非常に高い処理速度の記憶システムを提供するものである。異なるディスク駆動装置の主軸の回転を同期させる必要もなく、標準的な市販のディスク駆動装置と制御装置を使用することができる。本システムは、ディスク駆動装置全体の故障に対する裕度を有する。標準的なSCSIプロトコル指令を用いて、システムの全体的な動作を同期させる。

他の実施態様は願書の特許請求の範囲に含まれている。

例えば、第5図において、バス・アダプタ18は1台のEDC装置32を使用することができる。この場合、KBUSの64回線をそれぞれ32回線の2組に分け、その両方の組をマルチプレクサ30に対して接続する。データが格納される時、

動中、マルチプレクサ30は活動状態にあるがバンク42、44は不動作状態であることを保証し、さもなければマルチプレクサ30不動作でありバンク42、44はKBUSに対してビットを送ることができることを保証する。

EDC装置32に関して出入りするデータ・ビットの監視に基く論理回路48からインバータ46に対する入力を与えられる。回路48は連続する32ビットの組の状態に基いてフリップフロップのバンクをトリガーすることが適当である時を判定する。

各ディスク記憶装置により供されるKBUS回線数は、2つ以上または2つ以下のいずれにすることもできる。

4. (図面の簡単な説明)

第1図はコンピュータおよび関連する記憶システムを示すブロック図、第2図は第1図の1つの高速ディスク・サブシステムのブロック図、第3図は第2図のバス・アダプタの各部分を示すブロック図、第4A図および第4B図はデータの格納

マルチプレクサ30は、KBUSの1転送サイクルにおいて、最初にKBUSの1組の32回線を装置32の両方向WD/CD入力側に対して接続し、次いでKBUSの他方の組の32回線をWD/CD入力側に対して接続する。WD/CD入力側に対して接続された32回線はまた、それぞれ32個の8ビットのシフトレジスタ34の入力側に対して接続される。データ検索中、装置32はWD/CD出力における32の訂正されたデータ・ビットを32個のフリップフロップのバンク(LEF)40に対して送る。次に、シフトレジスタ34、36はそのバイトの第2のビットをEDC装置32に対してアンロードし、このEDC装置は更に32の訂正済みビットをフリップフロップのバンク(HEF)42に対して送出する。同時に、フリップフロップのバンク40はその32ビットをフリップフロップの第3のバンク(HEF)44に対してシフトする。この時、KBUSの取り引きが64ビット全てをコンピュータに対して送出することになる。インバータ46は、コンピュータがKBUSを駆

およびデータの検索操作を示すタイミング図、および第5図は第3図に代る回路図である。

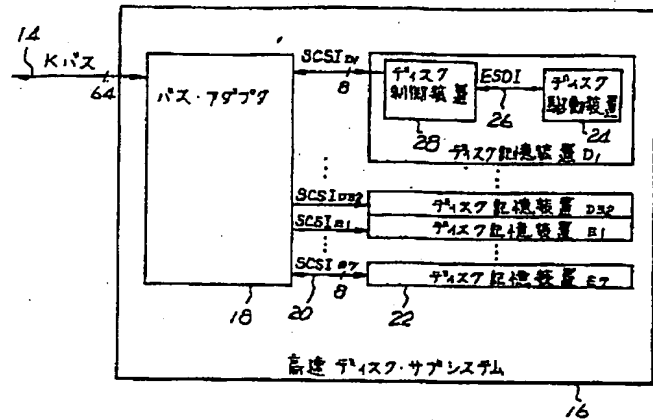
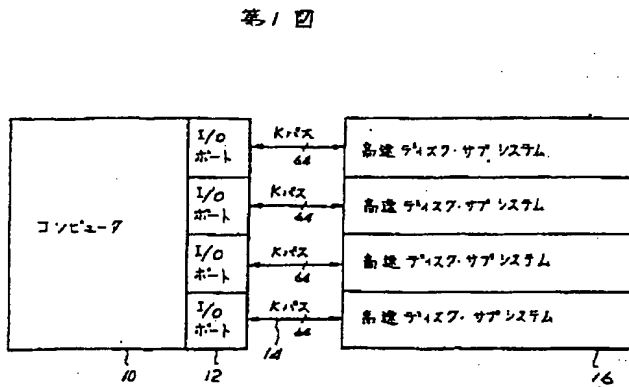
10…コンピュータ、12…I/Oポート、14…並列バス(KBUS)、16…高速ディスク・サブシステム、18…バス・アダプタ、20…並列バス(SCSIバス)、22…ディスク記憶装置、24…ディスク駆動装置、26…直列バス(ESDIバス)、28…ディスク制御装置、30…マルチプレクサ、32…EDC装置、34…シフトレジスタ、36…シフトレジスタ、40…フリップフロップのバンク(LEF)、42…フリップフロップのバンク(HEF)、44…フリップフロップのバンク(HEF)、46…インバータ、48…論理回路。

代理人 弁理士 湯 浅 恭 三



(外5名)

第2図



第3図

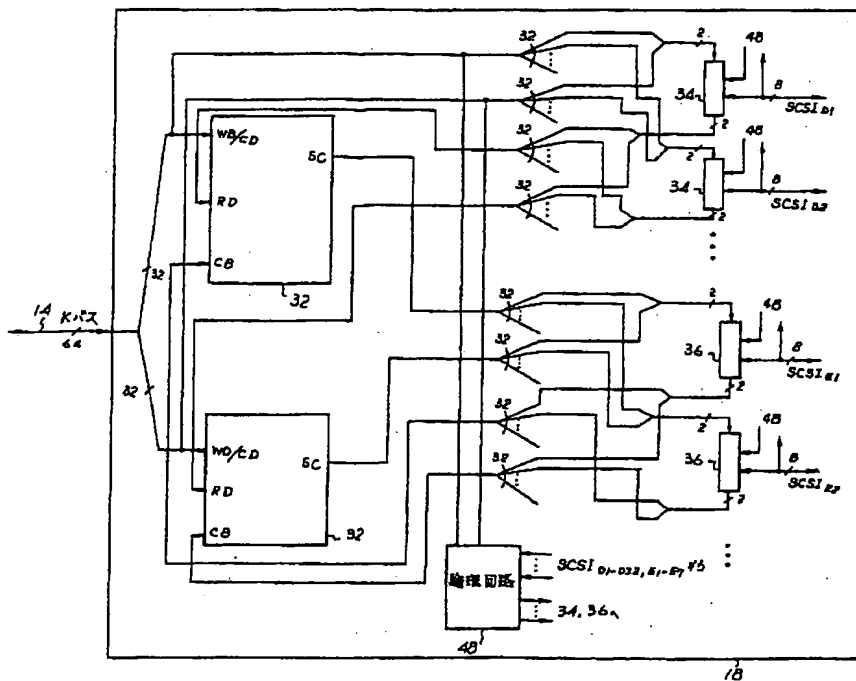
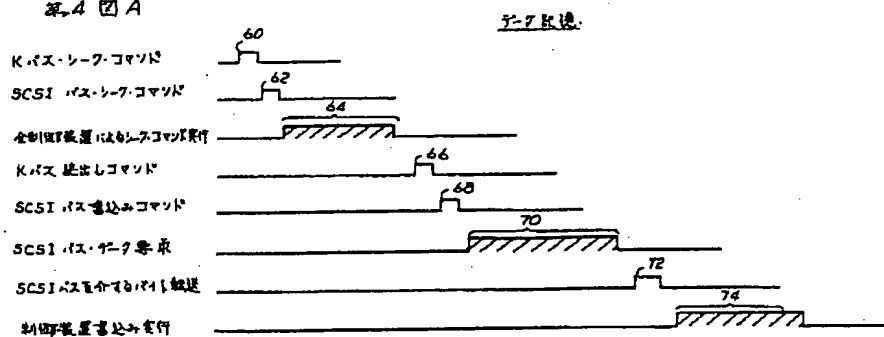
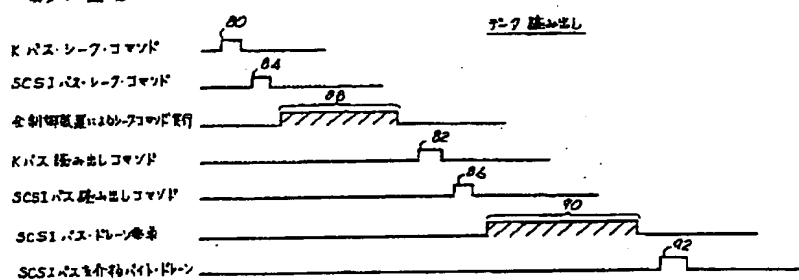


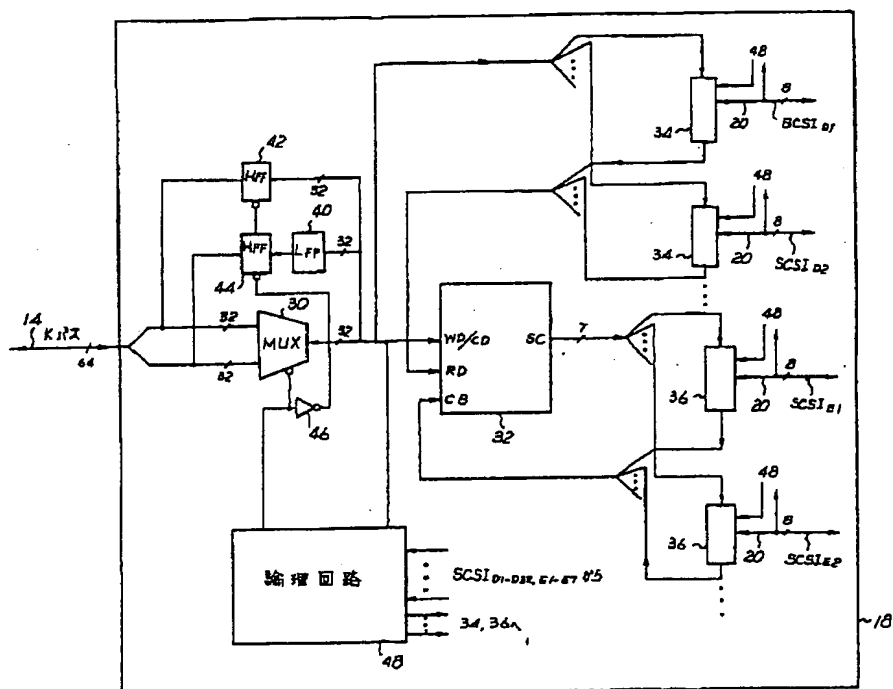
表 4 图 A



第 4 圖 B



第 5 回



【公報種別】特許法第17条の2の規定による補正の掲載
【部門区分】第6部門第4区分
【発行日】平成5年(1993)12月24日

【公開番号】特開昭62-24481
【公開日】昭和62年(1987)2月2日
【年通号数】公開特許公報62-245
【出願番号】特願昭61-105782
【国際特許分類第5版】

G11B 20/18 102 9074-5D
G06F 3/06 305 7165-5B

手 続 補 正 書

平成5年3月17日

特許庁長官 殿

1. 事件の表示

特願昭61-105782号

2. 発明の名称

記憶システム

3. 補正をする者

事件との関係 特許出願人
スィンキング マシンズ コーポレーション
(国籍) アメリカ合衆国

4. 代 理 人

〒107
東京都港区赤坂5丁目1番31号
第6セイコービル 3階
電 話 (03)3589-1201(代表)
(7748) 弁理士 谷 義 一

同 所

(8891) 弁理士 阿 部 和 夫

5. 補正命令の日付

自 発

6. 補正の対象

明細書および図面

7. 補正の内容

- (1) 明細書全文を別紙の通り補正する。
- (2) 図面全図を別紙の通り補正する。

(訂正) 明 細 書

1. 発明の名称

記憶システム

2. 特許請求の範囲

(1) 複数の複数ビットデータ部分をそれぞれ有するデジタルデータワードを記憶し、該データワードを検索要求に回答して並列データバスに連続して転送するシステムであって、

A. 複数のデータ記憶装置を有し、各データ記憶装置は機械的に駆動される媒体をそれぞれ有し、各データ記憶装置の媒体は互いに独立に機械的に駆動され、各データ記憶装置は各データワードに対して前記複数ビットデータ部分の個々の部分を記憶し、前記各データ記憶装置を用いてデータワードを記憶する複数データ記憶装置と、

B. 前記データワードが記憶され、前記機械的に駆動される媒体とは独立に、機械的に駆動される記憶媒体を含み、前記データワードに関連する

エラー訂正コードを記憶する少なくとも1つの訂正ビット記憶装置と、

C. 前記並列データバスと、前記データ記憶装置と、前記訂正ビット記憶装置とに接続されたアダプタであって、検索要求にตอบสนองして、(i) 前記データ記憶装置は選択されたデータワードの個々のデータ部分を同時に検索することができ、前記訂正ビット記憶装置は前記データ記憶装置による検索と同時に前記選択されたデータワードに関連するエラー訂正コードを検索することができ、(ii) 前記個々のデータ部分をアセンブルして前記選択されたデータワードにし、(iii) 前記関連するエラー訂正コードを用いて、前記選択された各データワードに関してエラー検査および訂正オペレーションを行い、前記選択されたデータワードが正しいか否かを判定し、判定した結果、正しくない場合は、個々に訂正されたデータワードを生成し、(iv) その後、前記選択されたデータワードを前記並列データバス上に伝送するアダプタとを備えたことを特徴とするシステム。

することができ、前記訂正ビット記憶装置による検索は、前記選択されたデータワードのデータ部分の前記データ記憶装置による検索と同時に行われ、(ii) 前記検索されたコード部分をエラー検出および訂正オペレーションにおいてエラー訂正コードとして用いるアダプタとを備えたことを特徴とするシステム。

(4) 特許請求の範囲第1項に記載のシステムにおいて、

前記アダプタは、

A. 前記データ記憶装置に対して、選択されたデータワードの個々のデータ部分を同時に検索させることができ、かつ、前記訂正ビット記憶装置に対して前記データ記憶装置による検索と同時に前記選択されたデータワードに関連するエラー訂正コードを検索させることができる記憶装置インタフェースと、

B. 該記憶装置インタフェースにより検査された前記個々のデータ部分をアセンブルして前記選択されたデータワードにするデータワードアセン

(2) 特許請求の範囲第1項に記載のシステムにおいて、

前記訂正ビット記憶装置は、データ記憶装置に障害が発生した場合、前記アダプタが前記選択されたデータワードに関してエラー検出および訂正オペレーションをパフォームしながら、前記障害が発生したデータ記憶装置上に記憶されたデータ部分に対する値を決定することができるようなエラー訂正ビットを記憶することを特徴とするシステム。

(3) 特許請求の範囲第1項に記載のシステムにおいて、

前記各エラー訂正コードは複数の複数ビットコード部分を含み、

前記システムは、

対応する数の訂正ビット記憶装置であって、前記コード部分のうちの1つの部分をそれぞれ記憶する訂正ビット記憶装置と、

(i) 前記訂正ビット記憶装置が前記選択されたデータワードに関連するコード部分を同時に検索

ブラと、

C. 該データワードアセンブラによりアセンブルされた各データワードに対して、関連するエラー訂正コードを用いて、前記選択されたデータワードに関してエラー検出および訂正オペレーションをパフォームし、選択され訂正されたデータワードを生成し、前記エラー訂正コード検査部分は前記選択されたデータワードが正しいと判定された場合、前記選択され訂正されたデータワードが前記選択されたデータワードに対応するエラー訂正コード検査部分と、

D. 前記選択され訂正されたデータワードであって、前記エラー訂正コード検査部分により生成されたデータワードを前記並列データバスに伝送するデータバスインタフェースとを含むことを特徴とするシステム。

(5) 特許請求の範囲第4項に記載のシステムにおいて、

各エラー訂正コードは複数の複数ビットコード部分を含み、

前記システムは、対応する数の訂正ビット記憶装置であって、前記コード部分の1つを記憶する訂正ビット記憶装置を含み、

前記記憶装置インタフェースは前記訂正ビット記憶装置に前記選択されたデータワードに関連するコード部分を同時に検索させることができ、

前記訂正ビット記憶装置による検索は前記選択されたデータワードのデータ部分の前記データ記憶装置による検索と同時に、

前記エラー訂正コード検査部分は検索されたコード部分をエラー検出および訂正オペレーションでエラー訂正コードとして用いることを特徴とするシステム。

(6) 特許請求の範囲第4項に記載のシステムにおいて、

前記記憶装置インタフェースは、予め定めた数の前記データ部分をそれぞれ備えた記憶ワードを各前記データ記憶装置から受信し、

前記データワードアセンブラは前記全記憶装置から受信された記憶ワードの個々のデータ部分か

た場合、前記アダプタが、前記各選択されたデータワードに関してエラー検出および訂正オペレーションをパフォームしながら、前記障害が生じたデータ記憶装置上に記憶されたデータ部分に対して値を決定するような訂正ビットを生成し前記ビット記憶装置に記憶することを特徴とするシステム。

(9) 特許請求の範囲第6項に記載のシステムにおいて、

前記アダプタは前記エラー訂正コードを複数の複数ビットコード部分として生成し、

前記システムは前記コード部分の1つをそれぞれ記憶する多数の訂正ビット記憶装置を含み、

前記アダプタは前記コード部分を個々の訂正ビット記憶装置に同時に伝送することを特徴とするシステム。

(10) 特許請求の範囲第6項に記載のシステムにおいて、

A. 前記データバスインタフェースは記憶要求とデータを前記並列バスからさらに受信し、

ら複数のデータワードをアSEMBルすることを特徴とするシステム。

(7) 特許請求の範囲第1項に記載のシステムにおいて、

前記データ記憶装置および前記訂正ビット記憶装置は、それぞれ、個々の記憶媒体に書き込む装置を含み、

前記アダプタは、記憶要求に応答して、データワードをさらに含み、

該データワードは、それぞれ、(i) 各データワードに対してエラー訂正コードを生成し、(ii) 各データワードを複数の複数ビット部分に分割し、(iii) 前記複数ビット部分およびエラー訂正コードを個々のデータ記憶装置および前記訂正ビット記憶装置に伝送し記憶する複数の複数ビット部分を含む

ことを特徴とするシステム。

(8) 特許請求の範囲第7項に記載のシステムにおいて、

前記アダプタは、データ記憶装置に障害が生じ

B. 前記記憶装置インタフェースはデータワード部分とエラー訂正コードを個々のデータ記憶装置および前記訂正ビット記憶装置に伝送し記憶し、

C. 前記アダプタは、

(i) 記憶される各データワードに対してエラー訂正コードを生成し、しかも、前記エラー訂正コードを、前記記憶装置インタフェースに転送し、該記憶装置インタフェースにより前記訂正ビット記憶装置に転送して記憶するエラー訂正コード生成装置と、

(ii) 各データワードを複数の複数ビット部分に分割し、前記データワードを前記記憶装置インタフェースに転送し、該記憶装置インタフェースにより前記データ記憶装置に転送して記憶するデータワード分割回路と

をさらに備えたことを特徴とするシステム。

(11) 特許請求の範囲第10項に記載のシステムにおいて、

前記記憶装置インタフェースは、前記各データ

記憶装置に、予め定めた数の前記データ部分をそれぞれ備えた記憶ワードを伝送し、

前記データワード分割回路は、記憶装置に伝送する記憶ワードを、各データワードのそれぞれ所定のデータ部分よりなるデータ集合から形成することを特徴とするシステム。

(12)特許請求の範囲第10項に記載のシステムにおいて、

前記各エラー訂正コードは複数の複数ビットコード部分を含み、

前記システムは、前記コード部分の1つをそれぞれ記憶する対応する数の訂正ビット記憶装置を含み、

前記エラー訂正コード生成装置は各エラー訂正コードを個々のコード部分に分割して前記記憶装置インタフェースに転送するエラー訂正コード分割回路を含み、

前記記憶装置インタフェースは前記コード部分を個々の訂正ビット記憶装置に転送し記憶することを特徴とするシステム。

書き込む装置を含む訂正ビット記憶装置と、

C. 並列データバスと、前記記憶装置と、前記訂正ビット記憶装置とに接続され、記憶要求にตอบสนองして、(i)各データワードに対してエラー訂正コードを生成し、(ii)各データワードを複数の複数ビット部分に分割し、(iii)前記複数ビット部分およびエラー訂正コードを個々のデータ記憶装置および前記訂正ビット記憶装置に伝送し記憶することを特徴とするシステム。

(14)特許請求の範囲第13項に記載のシステムにおいて、

前記アダプタは、データ記憶装置に障害が発生した場合、前記アダプタが、前記各選択されたデータワードに関して、エラー検出および訂正オペレーションをパフォーマンスしながら、前記障害が発生したデータ記憶装置に記憶されたデータ部分に対して値を決定するような訂正ビットを生成し前記ビット記憶装置に記憶することを特徴とするシステム。

(15)特許請求の範囲第13項に記載のシステムに

(13)ディジタルデータワードを記憶するシステムであって、各データワードは複数の複数ビットデータ部分を有し、前記システムが並列データバスを介して受信された記憶要求にตอบสนองして前記データワードを記憶するシステムにおいて、

A. 複数のデータ記憶装置を有し、各データ記憶装置は機械的に駆動される媒体をそれぞれ有し、各データ記憶装置の媒体は互いに独立に機械的に駆動され、各データ記憶装置は各データワードに対して前記複数ビットデータ部分の個々の部分を記憶し、前記各データ記憶装置を用いてデータワードを記憶する複数データ記憶装置であって、各データ記憶装置の個々の記憶媒体上にデータ部分を書き込む装置を含む複数データ記憶装置と、

B. 前記データワードが記憶された機械的に駆動された媒体とは独立に機械的に駆動された記憶媒体を含み、前記データワードに関連するエラー訂正コードを記憶する少なくとも1つの訂正ビット記憶装置であって、記憶媒体上に訂正ビットを

において、

前記アダプタは、前記エラー訂正コードを複数の複数ビットコード部分として生成し、

前記システムは、前記コード部分の1つをそれぞれ記憶する多数の訂正ビット記憶装置を含み、

前記アダプタは前記コード部分を個々の訂正ビット記憶装置に同時に転送することを特徴とするシステム。

(16)特許請求の範囲第13項に記載のシステムにおいて、

A. 記憶要求およびデータを前記並列バスから受信するデータバスインタフェースと、

B. データワード部分およびエラー訂正コードを個々の記憶装置および前記訂正ビット記憶装置に転送する記憶装置インタフェースと、

C. 記憶する各データワードに対してエラー訂正コードを生成し、前記エラー訂正コードを前記記憶装置インタフェースに伝送し、該記憶装置インタフェースにより前記訂正ビット記憶装置に転送して記憶するエラー訂正コード生成装置と、

D. 各データワードを複数の複数ビット部分に分割し、前記データワードを前記記憶装置インタフェースに伝送し、該記憶装置インタフェースにより前記データ記憶装置に伝送し記憶するデータ分割回路と

を含むことを特徴とするシステム。

(17)特許請求の範囲第16項に記載のシステムにおいて、

前記記憶装置インタフェースは、前記各データ記憶装置に、予め定めた数の前記データ部分をそれぞれ備えた記憶ワードを伝送し、

前記データワード分割回路は、記憶装置にそれぞれ伝送される記憶ワードを、各データワードのそれぞれ所定のデータ部分よりなるデータ集合から形成することを特徴とするシステム。

(18)特許請求の範囲第16項に記載のシステムにおいて、

各エラー訂正コードは複数の複数ビットコード部分を含み、

前記システムは、前記コード部分の1つをそれ

ぞれ記憶する対応する数の訂正ビット記憶装置を含み、

前記エラー訂正コード生成装置は生成された各エラー訂正コードを個々のコード部分に分割し、前記記憶装置インタフェースに転送し、

前記記憶装置インタフェースは前記コード部分を個々の訂正ビット記憶装置に転送し記憶することを特徴とするシステム。

(以下余白)

3. 発明の詳細な説明

〔産業上の利用分野〕

本発明は、磁気ディスクのような機械的に駆動される記憶媒体上にデジタルデータを記憶する記憶システムに関する。

〔従来の技術〕

典型的な磁気ディスク駆動装置は、いくつかの磁気ハードディスクが回転している単一のスピンドル上に取り付けられている。各ディスク面上の記憶空間は、同心円上のトラックにより編成されている。全てのディスク面上の対応するトラックにより、仮想のシリンドラが形成される。各ディスク面は読出し／書込みヘッドによりサーブされる。全ての読出し／書込みヘッドは、任意に選択されたシリンドラに一体となって移動させることができ、このシリンドラを構成するトラック上にデータを記憶するか、あるいはそのトラックからデータを検索することができる。

通常、磁気ディスク駆動装置は直列バスにより

ディスク制御装置に接続されており、このディスク制御装置は、読出し／書込みヘッドを制御して適正なトラックに移動させ、ディスクに格納されようとするか、あるいはディスクから検索されたばかりのデータをバッファリングし、データが適正なアドレスに格納されるか、あるいはこのアドレスから検索されることを保証し、かつデータに対して逐次エラー訂正を行い検出を行う。典型的には、このディスク制御装置は並列入出力バスをサーブする。この並列入出力バスは入出力ポートが例えば32ビット幅であるコンピュータに接続されている。この並列入出力バス上に転送される32ビットワードは、ディスク制御装置を介して渡され、直列に、ディスク駆動装置にロードされるか、あるいはこのディスク駆動装置から検索される。ディスク制御装置とディスク駆動装置との間の直列バスのスループットレートは、コンピュータとディスク制御装置の間の並列I/Oバスのスループットレートとおおよそ一致している。

利用可能な全記憶空間を拡張するため、複数のディスク駆動装置をデージーチェーンまたはスター構成で単一の制御装置にリンクすることができる。

〔課題を解決するための手段〕

本発明の全般的な1つの特徴は、データワードに対するエラー訂正ビットを生成し、このデータワードとは独立している訂正ビットを機械的に駆動される別個の媒体上に記憶することにある。

本発明の好ましい実施態様は、下記の特徴を含んでいる。各データワードの異なる部分は、異なるデータ記憶装置に記憶され、訂正ビットを記憶するための少なくとも1つの訂正ビット記憶装置がある。エラーは、任意の単一の記憶装置の障害に対してトレラントであるように訂正される。所定のデータワードに対するエラー訂正ビットのうちの種々のエラー訂正ビットは異なる訂正ビット記憶装置に記憶されている（所定のデータワードに対して2つのエラー訂正ビットが1つの記憶装

る。

本発明の好ましい実施態様は次の特徴を含む。1つのデータワードの全ての部分が転送実行可能になるまで、このデータワードの任意の部分を遅延させてバスに転送することにより、そのデータワードが検索される。各記憶装置は、要求されたデータワードのその部分が転送実行可能になるとドレーンコマンドを送信し、全ての記憶装置が待ち行列処理コマンドを発行した時にのみ、記憶装置がドレーンされる。並列のエラー訂正ビットが各データワードごとに生成され、関連するデータワードのどの部分も保持しない記憶装置上に、このエラー訂正ビットのうちの少なくとも1つが、記憶される。エラー訂正ビットはデータワードが読み出されるとそれに応じて読み出され、そのエラー訂正ビットを用いてデータワードの単一のビットエラーが訂正される。データワードを記憶する時、そのデータワードの異なる部分は、これらが全て記憶実行可能になるまで、異なる記憶装置に転送されない。各記憶装置は1つのバッファ

置に記憶されないのが好ましい）。エラー訂正ビットの集合は1つのデータワードを構成する全ビットよりも少ないビットのグループに対して同時に生成され、しかも、データワードビットのグループは次々に処理され、エラー訂正ビットの集合が連続して生成される。あるいはまた、異なるグループのデータワードのビットが同時に並列に処理されて、エラー訂正ビットの集合を生成する。例えば、各データワードは、それぞれ32ビットの2つのグループを構成する64ビットを持つことができ、7つの訂正ビットよりなる集合を32ビットの各グループ毎に生成することができる。データワードは、記憶および検索のため、並列データバスを介して連続して転送される。バスと記憶装置の間にアダプタが接続されており、このアダプタにより、このバスと記憶装置の間でデータワードの経路指定が行われる。

本発明の別の特徴は、非同期に機械的に駆動される異なる記憶装置の並列バス上に出力される各データワードの異なる部分を記憶することにあ

る。有し、読み出される連続データワードは、バッファが一杯にならないことを保証できるだけの速度で要求され、書き込まれる連続ワードはビットが空にならないことを保証できるだけの速度で転送される。各データワードの等しい数のビット（例えば、2ビット）が異なる記憶装置に記憶される。1つの記憶装置をサブする個々のバスはそれぞれSCSIプロトコルに準拠している。各データワードには33ビット以上（例えば、256ビット）が存在する。これらのデータワードは、バス線当り毎秒少なくとも4メガビットで並列バスに転送される。各記憶装置は磁気ディスク駆動装置である。

本発明は、非常に大きいビット幅の高速並列I/Oバスにサブすることができる非常に大きなスループットレートの記憶システムをサブする。標準的な市販のディスク駆動装置および制御装置は、異なるディスク駆動装置のス핀ドルの回転と同期をとることなく用いることができる。エラー訂正ビットおよびデータビットを独立して

記憶することにより、このシステムはどのディスク駆動装置全体の障害に対してもトレラントである。標準的なSCSIプロトコルコマンドを用いてシステムの全体的な動作を同期させる。

本発明の他の効果および特徴は、以下の好ましい実施態様の記述および特許請求の範囲から明らかになるであろう。

【実施例】

（構造およびオペレーション）

第1図を説明する。コンピュータ10は、（本願と同じ譲受人に譲渡された1983年5月31日出願の米国特許出願第499,474号に記載された種類の）コンピュータであり、4つの同じ入出力（I/O）ポート12を備えている。米国特許出願第499,474号の番号を付して本実施例の一部とする。各I/Oポート12は、64ビット幅の並列バス（KBUS）14を介して、対応する高速ディスクサブシステム16に接続されている。高速ディスクサブシステム16は、このKBUSを介して受信したデー

タを記憶し、かつ検索されたデータをこのKBUSに転送する。各KBUSは毎秒32メガバイト（バイト当り8ビット）のスループットレートを有し、毎秒128メガバイトの総スループットレートで4つのKBUS全ての上に同期してデータを転送することができる。

第2図を説明する。各ディスクサブシステム16はKBUSの1つに接続されたバスアダプタ18を有する。また、このバスアダプタ18は各々8ビット幅の39個の異なる並列バス（SCSIバス）20を介して、39個の異なるディスク記憶装置22に接続されている。各ディスク記憶装置22は、直列バス（ESDIバス）26を介してディスク制御装置28（例えば、Emulex社のチャンピオンモデル）に接続されたディスク駆動装置24（例えば、Control Data社の9415型）を含む。ディスク制御装置28はそのディスク記憶装置をサブするSCSIバスに接続されている。各SCSIバスは、米国商務省、規格標準局の規格刊行物第X3T9 2/8 2-2号、改定第14B号（1984年11月6日）に規定

された小型コンピュータシステムインタフェース規格に準拠してデータおよびコマンド（コマンドディスクリプタブロック-CDBの形式の）を送る。この規格刊行物の番号を付して実施例の一部とする。

39個のディスク記憶装置および39個のSCSIバスは、2つのグループで編成されている。ディスク記憶装置およびバスのうちの32個（D1ないしD32で示される）は、コンピュータ10からのデータの記憶および転送のため使用される。残りの7個のディスク記憶装置およびバス（E1ないしE7で示される）は、次に説明する方法でデータビットから取り出されたエラー訂正ビットを記憶し転送するため用いられ、エラー訂正のために用いられる。39個のディスク記憶装置およびSCSIバスは、あつかう情報の種別に区別があるだけで、同じものである。

各ESDIは、例えば、Magnetic Peripherals社の刊行物第77738076-D号（1984年）において規定された拡張小型装置インタフェースに準拠して

データおよびコマンドを逐次にする。

第3図を説明する。各バスアダプタ18において、KBUSの64本の線は各集合がそれぞれ32本の線によりなる2つの集合に分割される。これらの2つの集合はそれぞれ2つのエラー検出兼訂正装置32（それぞれ、例えば、1対のIntel社の8206型またはAMD社の2960型チップを備えている）の両方向WD/CD入力（書き込みデータまたは訂正されたデータ）端子に接続されている。また、各WD/CD入力端子に接続された32本の線は、それぞれ32個の8ビットシフトレジスタ34の入力端子に接続されている。各シフトレジスタはそれぞれ32本の線よりなる2つの集合からの1本の線に、その2ビットの入力が接続され、各シフトレジスタは2ビットを一時にシフトするように構成されている。

各エラー検出兼訂正装置32のWD/CD入力端子に入力される32個のデータビットの集合に対して、それぞれ、エラー検出兼訂正装置32は7つのエラー訂正ビットを並列に生成する。これらのエ

ラー訂正ビットは、関連する32個のデータビットを検索する間に、これら32個のデータビットのうちの任意のビットに生じる任意の単一ビットエラーを訂正することができるだけの情報を提供する。

各エラー検出兼訂正装置32の7つのエラー訂正ビットが7つのエラー訂正ビット線を介してSC出力端子から転送される。これらのビット線はレジスタ34のような7つの8ビットシフトレジスタ36の入力端子にそれぞれ接続されている。

KBUSの1つの転送サイクルが経過した後、各シフトレジスタ34はKBUSの線のうちの2つの線に出力されるデータビットと対応する2つのデータビットを含む。また、各シフトレジスタ36は対応するエラー訂正ビットのうちの2つのエラー訂正ビットを含む。KBUSの4つの転送サイクルが経過した後、各シフトレジスタ34は8つのデータビット（1データバイト）を含み、各シフトレジスタ36は8つのエラー訂正ビット（1エラー訂正バイト）を含む。その時、シフトレジス

のデータバイトの第1と第2ビットをEDC装置32のCB（検査ビット）入力端子にアンロードする。EDC装置32は、そのCB入力端子を介して受信されたエラー訂正ビットを用いて、そのRD入力端子で受信されたデータビットに現われる単一ビットエラーを検査しこれを訂正する。そして、EDC装置32はそのWD/CD出力端子の64個の訂正されたデータビットをKBUSに転送する。このプロセスは、シフトレジスタ34,36に保持されたバイトの8ビットが全てアンロードされるまで繰り返される。ついで、シフトレジスタは、SCSIバスからの新しいバイトにより再ロードされる。

各ディスク記憶装置28（第2図）は、ディスク駆動装置24上のデータの記憶および検索を次のように制御するように構成されている。

記憶オペレーションを開始するため、コンピュータ28は適正なコマンドディスクリプタブロック（CDB）をKBUS上に送信し、指定された数のバイト（ワードカウント）が、指定され

た34,36のバイトは、それぞれ、SCSIバスD1～D32およびE1～E7上にアンロードされる。以後、KBUSの4つの転送サイクル毎に、このプロセスが繰り返される。シフトレジスタ34,36のロードおよびアンロードのタイミングがとられ、毎秒32メガバイト（1つの線当り4メガバイト）でKBUS上に出力されるデータが、1つのSCSIバス当り毎秒1.5メガバイトのレートで、8ビット幅のSCSIバスのうちの32個のSCSIバスに転送される。逆に、1つのSCSIバス当り毎秒1.5メガバイトで出力されるデータは、毎秒32メガバイトでKBUSに転送される。

SCSIバスの各バイト転送サイクルにおいて、データがディスク駆動装置24から検索されると、1つの8ビットバイトがSCSIバス20を介して並列に各シフトレジスタ34,36にロードされる。次に、全てのシフトレジスタ34が共に並列にそのエラー訂正バイトの第1と第2ビットをEDC装置32のRD（読み取りデータ）入力端子にアンロードし、全てのシフトレジスタ36が並列にそ

たディスクアドレスから記憶されることを示す。このCDBは、バスアダプタ18によりSCSIバスを介してディスク制御装置28に渡される。ディスク制御装置28は必要な場合、ESDIバスを介してシークコマンドをディスク駆動装置に送信する。このシークコマンドにより、読み出し/書き込みヘッドを、CDBでアドレスが指定された記憶場所を含むシリンドラに移動させる。次に、ディスク制御装置28は記憶されるデータをSCSIバス上に転送するように要求することになる。この要求はKBUSを介してコンピュータに戻される。ついで、このデータは、（前に述べたように）KBUS上に渡され、バスアダプタによりSCSIバスに渡される。データビットが受信された時、これらビットは制御装置の14Kバイトのバッファに記憶される。各トラックはその容量が10Kバイトであるので、前記バッファは一時に全トラック相当以上のデータを保持することができる。ディスク制御装置28は、バッファが一杯になったときか（ワードカウントがバッファ容量を超えるか、あ

るいはこれと等しい場合)、あるいはバッファのバイト数がワードカウントと等しくなると直ちに、そのバッファをE S D Iバスに逐次アンロードすることを開始することになる。そして、データはトラック上の指定されたアドレスから格納される。ワードカウントがバッファ容量を超えた場合、データがバッファに8Kバイトしか残らないレベルまでドレーンされると直ちに、制御装置はコンピュータ10に対して別のデータを要求する要求を送信し、制御装置の古いデータがディスク駆動装置24にアンロードされると同時に、別のデータがバッファにロードされる。このプロセスを、ワードカウントが所定値に達するまで継続する。

同様に、検索オペレーションを開始するため、コンピュータ28は、指定されたディスクのアドレスから指定されたバイト数が検索されようとしていることを示すC D Bを送信する。制御装置により、必要な場合、最初に、ディスク駆動装置の読出し/書き込みヘッドは適正なトラックにシーク

／書き込みヘッドの下に現われない虞がある。従って、異なる制御装置が記憶コマンドを発行することができる時点に差があり、しかも検索されたデータを、異なる制御装置が転送実行可能になる時点に差があることになる。異なる制御装置を潜在的に位相がずれてオペレーションすることと、次の方法でバスアダプタとの間で同期して転送する必要とを妥協させる。

第4A図を説明する。データが記憶されると、コンピュータ10は最初にバスアダプタに特定のシリンダを命名するシークコマンド(60)を送信する。ついで、このバスアダプタは全ての制御装置にS C S Iバスを介して適正なC D Bを渡すことにより(62)、シークコマンドを同時通報して、全ての制御装置に同一のシリンダをシークさせる。制御装置が独立してシークコマンドを実行している間、コンピュータは他の関連のないオペレーションを行うことができる。

全ての制御装置がシークコマンドを実行したことを保証するに充分な期間(64)が経過した後、コ

し、指定されたアドレスからデータを読み出す。読み出されたデータはE S D Iバス上に逐次渡され、制御装置の直列バッファにロードされる。バッファがこのトラックのバイト数と等しいバイト数を含むと直ちに、制御装置はS C S Iバス上をコンピュータ10に要求を送信し、そのバッファからデータを取得する。このデータがアンロードされている間、制御装置はE S D Iバスを介してディスク駆動装置からデータを受け取り、そのデータを制御装置のバッファにロードし続ける。このプロセスは、バイト数がワードカウントと等しくなるまで継続される。

制御装置に接続された全てのS C S Iバスを介するデータの記憶および検索は、並列的に同時に起る。すなわち、バスアダプタの観点からは、関係するデータが単一のオペレーションで同時に全てのS C S Iバス上に渡される。異なるディスク駆動装置24は典型的には相互に位相がずれて回転しているので、異なるディスク駆動装置上の対応するトラック位置が、同時に、その各々の読出し

ンピュータ10は書き込みコマンド(シリンダ境界が交差しないようにする関連するディスクアドレスおよびワードカウントを含む)をバスアダプタに送信する(66)。このバスアダプタは、適正なC D BをS C S Iバスを介して全ての制御装置に送信することにより、全ての制御装置に書き込みコマンドを同時通報する(68)。そして、制御装置は、全て、格納されるデータをバスアダプタに要求する一連の要求(70)をS C S Iバス上にバイト単位で開始することになる。異なる制御装置からのデータは同期して出力されないのが典型的である。バスアダプタは、論理回路48を含み、この論理回路48は全てのS C S Iバスに接続され、要求が出されると、これを検出する。また、論理回路48はその制御出力端子が各シフトレジスタ34,36の制御入力端子に接続され、S C S Iバスを介して、シフトレジスタのロードおよびアンロードを制御する。

全ての制御装置がその要求を発行した時、バスアダプタは論理回路48から全てのシフトレジスタ

34,36 にアンロード信号を送信することにより、全ての要求を同時に満たし、シフトレジスタ 34,36 は全ての S I S C バス上にそのバイトを同時にアンロードする(72)。各制御装置はそのバッファが一杯になるまで(あるいは、そのバッファのバイト数がワードカウントと等しくなるまで)、データをバイト単位で要求し続けることになる。そして、各制御装置はそのデータをそのバッファからディスクに書き込むことを開始する(74)。

異なる制御装置は、個々のディスク駆動装置のシャフト位置の間の位相差に依存する非同期時点において、書き込みオペレーションを実行することになる。ワードカウントがバッファ容量を超えた場合、そのバッファに残っているバイト数が 8 K バイト未満になると直ちに、各制御装置はバスアダプタからさらにデータを要求することになる。これらの制御装置によるこのデータ要求は、バスアダプタに非同期的に到達することになる。さらにバイトを送信することによりこれら要求を満た

も、ディスク駆動装置がディスクからバイトを読み出すことにより、各制御装置は読み出しコマンドの実行を開始し、読み出されたバイトを E S D I バス上に逐次転送し、制御装置のバッファにロードさせる。異なる制御装置は、データ読み出しを非同期に実行するのが典型的である。セクター分の読み出しデータがバッファに入力された時、制御装置によりバッファドレーンが要求される(90)が、データ読み出しを継続するとともに、バッファを満たし続けることになる。全ての制御装置がバッファドレーン実行待機中であることをバスアダプタの論理回路が検出すると、この論理回路はこれらバッファのドレーンオペレーションを開始する。その後、バスアダプタは、最初にドレーンを要求する制御装置により要求された速度で、制御装置のバッファからバイトをドレーンし続けることになり、従って、最初のバッファが決して一杯にならないことを保証する。読み出しオペレーションは、各ディスク駆動装置から読み出されたバイト数がワードカウントと等しくなるま

す前に、全ての要求が受信されるまで、バスアダプタはその論理回路 48 を介して待機することになる。実際には、最後に、データをさらに要求する制御装置のバッファが 8 K バイトのレベルにドレーンされるまで、バスアダプタは待機する。一方、最初に、データをさらに要求した制御装置のバッファは、8 K バイトレベル未満だけ空にされる。その後、バスアダプタは最初に要求する制御装置がデータバイトを要求すると、直ちに、データバイトを送信することになる。よってバスアダプタは、最初に要求する制御装置のバッファが決して完全に空にならないことを保証する。このプロセスは、各制御装置に転送されるデータバイト数がワードカウントに達するまで継続する。

第 4 B 図を説明する。データが検索されると、コンピュータ 10 は読み出しコマンド(82)の後にシークコマンド(80)を送信する。両コマンドはバスアダプタによってデータ記憶オペレーションと同様の方法で制御装置に同時通報される(84,86)。各制御装置はシークコマンド(88)を実行する。しか

で継続される。

各高速ディスクサブシステムのディスク記憶装置 22 のうちのどれかに障害が発生すると、E D C 装置は障害が発生した装置から受信した誤ったビットを訂正することが可能である。そして、ディスクサブシステムのオペレーションを妨げることなく、障害が発生した装置全体を正常な装置と交換することができる。

本発明は、スループットレートが非常に高い記憶システムであって、非常に大きいビット幅の高速並列 I/O バスをサージすることができる記憶システムを提供する。異なるディスク駆動装置のスピンドルの回転を同期させる必要もなく、標準的な市販のディスク駆動装置と制御装置を使用することができる。本システムは、1つのディスク駆動装置全体の障害に対してトレラントである。標準的な S C S I プロトコルコマンドを用いて、システムの全体のオペレーションを同期させる。

他の実施態様は特許請求の範囲に含まれる。

第5図を説明する。例えば、バスアダプタ18は1台のEDC装置32を使用することができる。この場合、KBUSの64本の線をそれぞれ32本の線の2つの集合に分け、その両方の集合をマルチプレクサ30に接続する。データが記憶される時、マルチプレクサ30は、KBUSの単一の転送サイクルの間に、最初に、KBUSの32本の線の1つの集合をEDC装置32の両方向WD/CD入力端子に接続し、ついでKBUSの32本の線の他の1つの集合をWD/CD入力端子に接続する。また、WD/CD入力端子に接続された32本の線は、それぞれ、32個のBビットシフトレジスタ34の入力端子に接続される。データ検索中、EDC装置32はWD/CD出力端子の32個の訂正されたデータビットを32個のフリップフロップのバンク(LFF)40に転送する。次に、シフトレジスタ34,36はそのバイトの第2のビットをEDC装置32にアンロードし、このEDC装置32は32個の訂正済みビットをフリップフロップのバンク(HFF)42に転送する。同時に、フリップフロップのバンク

40はその32ビットをフリップフロップの第3のバンク(HFF)44にシフトする。この時、KBUSのトランザクションが64ビット全てをコンピュータに転送することになる。インバータ46は次のことを保証する。すなわち、コンピュータがKBUSを駆動している間、マルチプレクサ30は活動状態にあるが、バンク42,44は非活動状態であり、他方、コンピュータがKBUSを駆動していない間は、マルチプレクサ30は非活動状態にあり、バンク42,44はKBUSにビットを転送することができることを保証する。

EDC装置32と受け渡しをするデータビットの監視に基づき、インバータ46に論理回路48から入力される。論理回路48は、連続する32ビットの集合の態様に基づいて、フリップフロップのバンクのトリガーを何時行うのが適正かを判定する。

各ディスク記憶装置によりサーブされるKBUSの線数は、3本以上または3本以下のいずれかにすることができる。

4. 図面の簡単な説明

第1図はコンピュータおよび関連する記憶システムを示すブロック図、

第2図は第1図に示す1つの高速ディスクサブシステムを示すブロック図、

第3図は第2図のバスアダプタの各部分を示すブロック図、

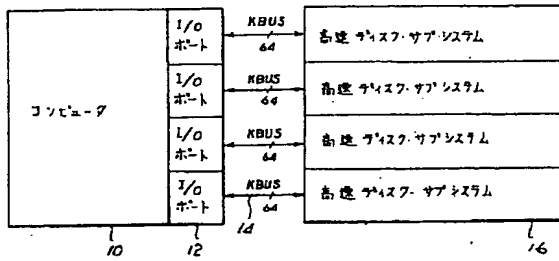
第4A図および第4B図はデータの記憶およびデータの検索オペレーションを示すタイミング図、

第5図は第3図に代わる回路図である。

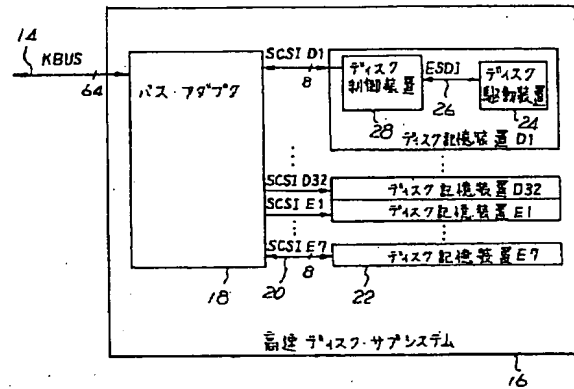
- 10…コンピュータ、
- 12…I/Oポート、
- 14…並列バス(KBUS)、
- 16…高速ディスクサブシステム、
- 18…バスアダプタ、
- 20…並列バス(SCSIバス)、
- 22…ディスク記憶装置、
- 24…ディスク駆動装置、

- 26…直列バス(ESDIバス)、
- 28…ディスク制御装置、
- 30…マルチプレクサ、
- 32…EDC装置、
- 34…シフトレジスタ、
- 36…シフトレジスタ、
- 40…フリップフロップのバンク(LFF)、
- 42…フリップフロップのバンク(HFF)、
- 44…フリップフロップのバンク(HFF)、
- 46…インバータ、
- 48…論理回路。

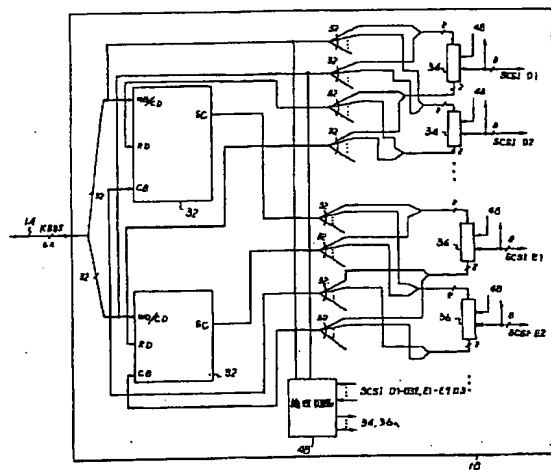
第1図



第2図

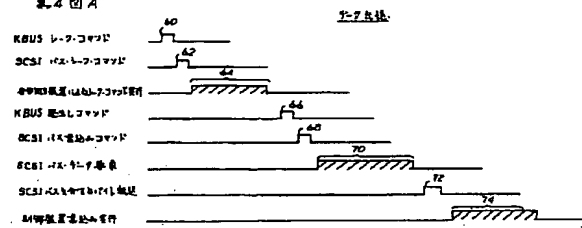


第3図



(補正図面)

第4図A



第4図B

